

## CUANTIFICACION VECTORIAL EN CODIFICACION DE VOZ POR EXCITACION MULTIPULSO

A.P. Honold Ripoll, A. Moreno Bilbao, J.A. Rodríguez Fonollosa

Dpto. Teoría de la Señal y Comunicaciones  
Universidad Politécnica de Cataluña

### ABSTRACT

In this paper is made a comparison between some quantizers applied in a Multipulse Speech coder. Vector quantizer is compared against scalar quantization in the LPC parameter in the short predictor. Adaptive Multistage Vector Quantizer gives the best results. Long predictor coefficient obtained either in open or closed loop is quantized and results are compared. Finally, excitation signal is quantized scalar and vectorially and results are compared. Any of the different VQ schemes tested gives a better performance in SNR terms than the scalar scheme

### 1.INTRODUCCION

En los últimos años se han realizado muchos avances en la codificación de voz a media y baja velocidad por medio de los denominados codificadores híbridos. Estos codificadores obtienen la señal sintetizada por medio del filtrado de una secuencia obtenida en un etapa de análisis de forma que minimice un criterio de error perceptual entre la señal original y la sintetizada. La determinación de la secuencia se realiza mediante un bucle de análisis por síntesis.

Dos son básicamente los codificadores encuadrados en este grupo: codificador multipulso [1], y CELP [2] y sus distintas variantes.

En la codificación multipulso, la secuencia de excitación está formada por pulsos cuyas amplitudes y posiciones deben determinarse. Los codificadores CELP determinan la secuencia óptima entre un conjunto de secuencias previamente establecidas.

La codificación de voz por medio de codificadores por excitación multipulso permite obtener una alta calidad a velocidades comprendidas entre 8kbps y 10kbps.

Para lograr esta calidad es necesario cuantificar convenientemente todos los parámetros necesarios para generar la señal sintetizada. En esta comunicación se hace un estudio comparativo sobre la aplicación de técnicas de cuantificación vectorial (VQ) y cuantificación vectorial adaptativa (AVQ) en sistemas de codificación por excitación multipulso.

Esta comunicación está estructurada de la siguiente forma: en el próximo apartado se describen los parámetros y métodos de cálculo utilizados en este trabajo para realizar el codificador multipulso sobre el que se han realizado las pruebas. En el apartado III, se describen los cuantificadores que se han utilizado para realizar las comparaciones. En el apartado IV se presentan los resultados y en el último apartado las conclusiones.

## II CODIFICADOR MULTIPULSO

Los elementos que componen el codificador multipulso son: el predictor corto, el predictor largo o predictor de pitch y la excitación.

En la etapa de análisis, la señal de voz es muestreada a 8 KHz y dividida en tramas de 25 mseg por una ventana de Hamming. Esta segmentación se realiza cada 20 mseg. Para cada trama se obtiene el filtro de síntesis de orden 10 por técnicas de predicción lineal (método de autocorrelación).

La excitación multipulso consta de cuatro pulsos que se actualizan cada 5 mseg. El factor de ponderación del filtro perceptual ha sido  $\gamma=0.5$ . La amplitud y posición de cada pulso se calcula en forma iterativa. En cada iteración se calcula una nueva posición y amplitud dejando fijas las posiciones de los pulsos hallados previamente. Las amplitudes de los pulsos pueden optimizarse en cada iteración o únicamente en la última [3]. En los resultados mostrados en este trabajo, las amplitudes son reoptimizadas en cada iteración.

El predictor de pitch (predictor largo) consta de un coeficiente que puede ser calculado en bucle abierto o en bucle cerrado [3] cada 5 mseg. El método de determinación del retardo y el valor del coeficiente en bucle abierto ha sido por minimización del residuo.

## III. CUANTIFICACION

Las técnicas de cuantificación vectorial han sido aplicadas sobre los parámetros LPC y sobre la excitación.

La cuantificación de los parámetros LPC se ha realizado con dos sistemas : Cuantificación Vectorial Multietapa MSVQ y Cuantificación Vectorial Multietapa Adaptativa AMSVQ [4]. AMSVQ utiliza varios cuantificadores vectoriales en cascada y el error de cuantificación es utilizado para adaptar la primera etapa mediante el algoritmo LMS.

El efecto de la cuantificación del predictor largo se estudió con un coeficiente en bucle abierto y un coeficiente en bucle cerrado. Se ha puesto especial atención en el caso de un coeficiente por razones de velocidad de señalización. Para cada método de cálculo se diseñó un cuantificador escalar específico.

Se han estudiado distintos esquemas de cuantificación ganancia-forma [5] para la excitación y se han comparado con la cuantificación escalar :

- Cuantificación vectorial de una etapa más signo
- Cuantificación vectorial de una etapa y extracción del máximo
- Cuantificación vectorial multietapa MSV
- Cuantificación vectorial multietapa adaptativa AMSVQ ( $\mu$  desde 0.01 a 0.1)

El estudio se ha realizado para una velocidad de señalización alrededor de los 9.6 Kbps. En la TABLA I se muestran los bits asignados a cada etapa, los periodos de actualización de los parámetros y la velocidad de señalización correspondiente. Una velocidad de señalización menor requeriría variar además los periodos de actualización de los distintos parámetros.

Parámetros	Bits	Actualización (ms)	Velocidad (b/s)
Predictor corto	19-36	20	950-1800
Predictor largo			
Retardo	7	5 - 10	700-1400
Amplitud	3	5	600
Excitación			
Ganancia	5	5	1000
Amplitudes	9-12	5	1800-2400
Posiciones	17	5	3400

TABLA I- Asignación de bits para cada etapa del codificador

## IV RESULTADOS

La base de datos de entrenamiento está constituida por 80 frases de 14 locutores diferentes y de aproximadamente 2seg de duración cada una. En el diseño de los cuantificadores vectoriales se ha asegurado que cada vector fuera diseñado con un mínimo de 50 vectores de entrenamiento. La evaluación de los codificadores se ha realizado con 12 señales de test pronunciadas por locutores que no pertenecen a la base de diseño.

La SNRSEG esta calculada en tramas de 160 muestras con un umbral de potencia de 512.

El diseño de los cuantificadores está dividido en tres fases, predictor corto, predictor largo y excitación. Los efectos de la cuantificación en cada etapa se muestran con los de las anteriores cuantificadas.

### a) Cuantificación del predictor corto.

En la cuantificación del predictor corto se utilizaron los coeficientes LAR. Para la cuantificación de estos coeficientes se asignaron entre 24 y 30 bits. Para aprovechar las dependencias entre coeficientes, se deben utilizar estructuras de coste reducido al no ser operativo un cuantificador de búsqueda completa de tamaño  $\geq 2^4$  vectores.

Se estudiaron distintas alternativas en cuanto a número de etapas y factor de adaptación de la primera etapa. Se comprobó que una cuantificación multietapa adaptativa permite cuantificar los coeficientes con 27 bits, lo que lleva a una velocidad de 1350 b/s, mientras que otros esquemas escalares necesitan 1800 b/s proporcionando una calidad similar.

En la TABLA II se muestran los resultados de SNR y SNRseg para un diseño multietapa con (3,8,8,8)bits por etapa y factores de adaptación  $\mu=0$  (sin adaptar),  $\mu=0.07$  y  $0.1$ . En la misma tabla se muestran los resultados de una configuración de 3 etapas y 19 bits (3,8,8 bits por etapa.). Los resultados sin cuantificar son SNR=15.48 dB y SNRSEG=16.14 dB.

	Esquemas del		cuantificador del		predictor corto	
	MSVQ( $\mu=0$ )		AMSVQ ( $\mu=0.07$ )		AMSVQ ( $\mu=0.1$ )	
Número de etapas del cuantificador	SNR (dB)	SNRSEG (dB)	SNR (dB)	SNRSEG (dB)	SNR (dB)	SNRSEG (dB)
4 etapas (3,8,8,8)	14.46	15.34	14.66	15.47	14.75	15.44
3 etapas (3,8,8)	13.90	14.87	14.05	14.97	14.16	14.93

TABLA II- Comparación de los resultados obtenidos al cuantificar el predictor corto mediante un cuantificador multietapa sin adaptar MSVQ y adaptando AMSVQ.

Los resultados muestran una clara ventaja de AMSVQ sobre MSVQ. Las pérdidas para el cuantificador AMSVQ de cuatro etapas son de 0.7 dB. Pruebas auditivas muestran que las diferencias apenas son apreciables. Para el esquema de tres etapas, las pérdidas son 0.5 dB superiores y según las audiciones realizadas, la degradación en la señal sintetizada ya se puede apreciar.

Para testear los efectos de la cuantificación en las sucesivas etapas del sistema se seleccionó un AMSVQ de cuatro etapas con (3,8,8,8) bits cada una.

### b) Cuantificación del predictor largo

En el predictor largo hay dos parámetros: el retardo y la amplitud. El cálculo puede realizarse en bucle abierto o cerrado como ya se ha comentado. En el caso de cálculo en bucle abierto, la actualización del retardo se puede realizar a un periodo doble que el de la amplitud. Este caso lo denominamos "trama mixta". El sistema es más sensible a una actualización cada 10 ms del valor de la amplitud que al valor del retardo. Los valores de retardo pueden variar

entre  $M_{\min} \leq M \leq M_{\max}$  con  $M_{\max} = M_{\min} + 127$  y  $M_{\min} = 16$  para el cálculo en bucle abierto y  $M_{\min} = 40$  para el cálculo en bucle cerrado. El valor del retardo se cuantifica con 7 bits.

Al valor de la amplitud se asignaron 3 bits. Para cuantificar las amplitudes se realizó un histograma para cada método de cálculo de sus valores sobre todo el conjunto de tramas sonoras de la base de datos. Los niveles del cuantificador se establecieron por el algoritmo LBG [6]. Cabe mencionar que, debido a la distribución de valores de amplitud del predictor largo calculado en bucle cerrado, el sistema es sensible (en pruebas auditivas aunque no en SNR) a la elección del margen dinámico del cuantificador previo a la aplicación del algoritmo LBG.

	Bucle abierto		Trama mixta		Bucle cerrado	
	SNR (dB)	SNRSEG (dB)	SNR (dB)	SNRSEG (dB)	SNR (dB)	SNRSEG (dB)
Sin cuantificar	15,48	16,14	14,65	15,32	15,66	16,64
Cuantificado	14,88	15,44	14,66	15,29	14,82	15,81

TABLA III. Resultados de la cuantificación del predictor corto

En la TABLA III se muestran los resultados para cada método sin cuantificar y cuantificando el predictor corto y el predictor largo. Puede apreciarse que, aunque el cálculo en bucle cerrado da mejores resultados sin cuantificar, el efecto de la cuantificación hace que el sistema sea inferior en calidad.

A partir de este momento se elegirán como valores de cuantificación los del predictor largo en bucle abierto a efectos de comparar los métodos de cuantificación de la excitación.

#### c) Cuantificación de la excitación.

La codificación de la excitación consta de dos partes: Posición de los pulsos en cada trama y cuantificación de sus amplitudes. Las posiciones se codificaron con 17 bits por trama según el algoritmo descrito en [7].

Para diseñar el cuantificador de las amplitudes se amplió la base de datos de entrenamiento con varias frases de cuatro locutores adicionales, para conseguir 50000 tramas de excitación (250 s de señal a los que se han separado los silencios iniciales y finales de cada frase). Se utilizó un cuantificador ganancia\_forma. La ganancia, obtenida como la amplitud máxima de los valores absolutos de los pulsos de la trama, fue cuantificada con un cuantificador Ley A de 5 bits. Para la forma se probaron los siguientes cuantificadores:

1. Cuantificador escalar de 3bits por pulso, diseñado por el algoritmo LBG.
2. Cuantificador vectorial con normalización de signo. 4 bits para el signo de cada pulso y un VQ de 8bits para cuantificar amplitudes en valor absoluto.
3. Cuantificación vectorial y extracción del pulso máximo. Un VQ de 9 bits para los tres pulsos de amplitud menor, 2 bits para la posición relativa del pulso de amplitud mayor y un bit para el signo. La amplitud de este pulso es próxima a uno.
4. Cuantificación vectorial multietapa MSVQ. Debido al número de bits tan reducido, únicamente se probaron esquemas con dos etapas, variando entre 9 y 12 bits.
5. Cuantificación vectorial adaptativa. Se probaron esquemas de dos etapas con  $\mu$  variando de 0.01 a 0.1 y asignando entre 9 y 12 bits.

Corto y largo cuantificados		Esquemas de cuantificación de excitación					
		Escalar		VQ +signo		VQ +máximo	
SNR (dB)	SNRSEG (dB)	SNR (dB)	SNRSEG (dB)	SNR (dB)	SNRSEG (dB)	SNR (dB)	SNRSEG (dB)
14.88	15.44	14.54	14.95	14.67	15.03	14.59	14.98

TABLA III Resultados con los distintos esquemas de cuantificación de la excitación

MSVQ (8,4)		MSVQ (4,7)		MSVQ (4,6)		MSVQ (4,5)	
SNR (dB)	SNRSEG (dB)	SNR (dB)	SNRSEG (dB)	SNR (dB)	SNRSEG (dB)	SNR (dB)	SNRSEG (dB)
14.57	14.91	14.63	14.63	14.25	14.39	13.97	14.18

TABLA IV Resultados con cuantificación multitapa para 12,11,10 y 9 bits

En las TABLAS III y IV se muestran los resultados para los distintos cuantificadores de la excitación con todas las etapas cuantificadas. Puede apreciarse que para el mismo número de bits, la cuantificación vectorial mejora a los resultados obtenidos con cuantificación escalar. El esquema que da mejores resultados es el de VQ con normalización de signo.

El esquema multitapa también ha sido utilizado para variar el número de bits asignados a la excitación. En la tabla IV se muestran los mejores resultados obtenidos para las distintas asignaciones de bits en cada etapa.

Los esquemas multiépapa adaptativos no mejoran, o muy ligeramente, la SNR obtenida sin adaptar.

## V. CONCLUSIONES

En este trabajo se ha realizado una comparación de métodos de cuantificación aplicados al codificador multipulso, con objeto de estudiar las influencias sobre un sistema que trabajaría alrededor de los 9.6Kbps.

Se ha comprobado que la cuantificación vectorial adaptativa AMSVQ es una buena elección para la cuantificación de los coeficientes del predictor corto. También se ha estudiado el efecto de la cuantificación sobre el predictor largo concluyéndose que la mejora aportada por el cálculo en bucle cerrado no compensa la pérdida de la cuantificación siendo más recomendable y no sólo por carga computacional, el cálculo en bucle abierto. En cuanto a la cuantificación de la excitación se ha comprobado que los sistemas de cuantificación vectorial dan un comportamiento mejor que la cuantificación escalar. La elección de VQ y normalización de signo sería una buena elección tanto por su calidad como por su relativamente bajo coste computacional.

## VI REFERENCIAS

- [1] Atal, Remde "A New Model of LPC excitation for producing natural-sounding at low bit rates" Proc. IEEE ICASSP'82
- [2] Schroeder, Atal "Code excited linear prediction (CELP): High quality speech at very low bit rates" Proc. IEEE ICASSP'85
- [3] Singhal, Atal "Amplitude Optimization and Pitch Prediction in Multipulse Coders" IEEE Transactions on ASSP. vol. 37, March 1989
- [4] Rodríguez Fonollosa "Cuantificación vectorial adaptativa aplicada a la codificación de voz." Tesis Doctoral. Universidad Politécnica de Cataluña, 1989
- [5] Sabin, Gray "Product code vector quantizers for waveform and voice coding" IEEE Transactions on ASSP. vol. 32, March 1984
- [6] Linde, Buzo, Gray "An algorithm for vector quantizer design" IEEE Transactions on Communications. vol. 37, March 1980.
- [7] Kroon, Deprettere. "A class of analysis by synthesis predictive coders for high quality speech coding at rates between 4.8 and 16 Kbps" IEEE Journal on selected areas, vol.6, n.2 1988.